DOCUMENT RESUME

ED 154 035 TM 007 112

AUTHOR Livingston, Samuel A.
TITLE Reliability of Tests Used to Make Pass/Fail Decisions: Answering the Right Questions.
PUB DATE Mar 78
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Toronto, Ontario, Canada, March, 1978); Appendices of marginal legibility

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Cutting Scores; Decision Making; Efficiency; *Mastery Tests; Measurement; *Pass Fail Grading; Standard Error of Measurement; *Test Reliability; True Scores
IDENTIFIERS *Estimation; *Lords Method 20

ABSTRACT

The traditional reliability coefficient and standard error of measurement are not adequate measures of reliability for tests used to make pass/fail decisions. Answering the important reliability questions requires estimation of the joint distribution of true and observed scores. Lord's "Method 20" estimates this distribution without the deficiencies of other methods. New output formats condense the estimated distribution into readily usable information, including a 2 x 2 contingency table, conditional true-score distributions, and an index of decision-making efficiency. Examples are appended. (Author/CTM)

ED154035

Presented at the NCME Annual Meeting:
Toronto, Ontario; March, 1978

Reliability of tests used to make pass/fail decisions: Answering the right questions.

SAMUEL A. LIVINGSTON
Center for Occupational and Professional Assessment
Educational Testing Service

The traditional reliability coefficient and standard error of measurement are not adequate measures of reliability for tests used to make pass/fail decisions. Answering the important reliability questions requires estimation of the joint distribution of true and observed scores. Lord's "Method 20" estimates this distribution without the deficiencies of other methods. New output formats condense the estimated distribution into readily usable information, including a 2 x 2 contingency table, conditional true-score distributions, and an index of decision-making efficiency.

TM007 112

Printed in U.S.A.

Reliability of tests used to make pass/fail decisions: Answering the right questions

Samuel A. Livingston
Center for Occupational and Professional Assessment
Educational Testing Service

As the title implies, this paper is about reliability - the reliability of a particular type of test. The tests I am concerned about are tests that have the following three characteristics: First, the test is made up of items scored as right or wrong. Second, the test score will be used as the basis for a pass/fail decision. Third, the variable measured by the test is continuous; a person's true score can be anywhere on the continuum. Notice that I am not talking about a situation in which the test scores are supposed to reflect a true dichotomy. In the kind of testing I am concerned about, the terms "mastery" and "non-mastery" are somewhat misleading; it makes more sense to talk about "minimally acceptable performance". It is easy to think of examples of tests of this type: A final exam for a course, a licensing exam for a profession, a screening test for a remedial program, and so on.

The traditional measures of test reliability are the reliability coefficient and the standard error of measurement. What do these traditional measures really tell us? The reliability coefficient answers the question: "What is the estimated correlation between two equivalent forms of the test, in the group of persons taking the test?" The standard error of measurement answers the question: "What is the average size of the errors of measurement, in the group of persons taking the test?" But

these are not the important questions. What neither the reliability coefficient nor the overall standard error of measurement tells us is how much confidence we can have in the pass/fail decisions we are making on the basis of the test.

What kind of questions should we be asking? The important questions of reliability have to do with the pass/fail decisions and the relationship of these decisions to the true scores of the persons taking the test. (By "true score", I mean the person's average score over the universe of all possible tests.) Some of the questions we should be asking are these:

1. Of those persons who passed the test, how many would have passed if the test had been perfectly reliable? Of those who failed, how many would have failed if the test had been perfectly reliable?
2. What is the distribution of true scores in the group of persons who passed the test? What is the distribution of true scores in the group of persons who failed the test?
3. What is the decision-making efficiency of the test, as compared with that of a perfectly reliable test?

To answer questions like these, we need to estimate the joint distribution of true scores and observed scores. This is not a simple problem. Fred Lord worked on it for several years, and he developed several methods of estimating this joint distribution. The most advanced of these was published in 1969. It is called simply "Method 20". The purpose of this paper is to describe briefly some of the characteristics of this procedure and to show how it can be used to answer the kinds of reliability questions we should be asking.

Method 20 assumes that the conditional distribution of observed scores, for persons with a given true score, is a "compound binomial distribution". This assumption is less restrictive than the assumption of a binomial distribution. Strictly speaking, the binomial assumption would be appropriate only if all test items were equally difficult or if each person taking the test received a separate random sample of items. (Actually, Method 20 can also be used with the binomial assumption, and the results turn out not to be very different.)

Method 20 is somewhat unusual in that the mathematical model for the true-score distribution does not have a fixed number of parameters. In fact, the number of parameters in the model is limited only by the number of items on the test. This feature makes the model very flexible. Even if the distribution of true scores happens to have some unusual shape, Method 20 can estimate it. Other methods - for example those based on the beta distribution - cannot estimate a bimodal distribution of true scores.

Method 20 does have some limitations. The test must be unspeeded, and there must be no correction for guessing. Ideally, every person should have a chance to answer every item, and the person's score must be simply the number of items answered correctly. The only limit on the length of the test is a practical one; the existing computer program will not accept a test of more than 100 items. Also, a very short test limits the complexity of the estimated true-score distribution. But even as few as six items will allow for a bimodal distribution with unequal modes.

The required input for Method 20 consists of the observed score distribution and the variance of item dificulty values. The catch is that this information has to come from a large sample - preferably 1000 or more persons. However, the input does not have to come from a test

of the same length as the one for which we want the estimates. That is, we can use data from a 20-item test to get an estimated distribution for a 50-item test, or a 10-item test, or a test of any length we want to specify. This feature is especially useful in deciding how long a test needs to be.

Since I am neither a mathematician nor a computer programmer, I won't attempt to explain either the mathematical solution or the computer programming of Method 20. Both were published in 1969. The mathematical solution is in an article by Lord in Psychometrika; the computer program is in an article by Wingersky and others in Educational and Psychological Measurement.

What is new is the collection of output formats we at ETS have recently developed, and the computer program that produces them. These formats take the estimated joint distribution of true and observed scores and condense it down to a readable, usable form, so that the director of a testing program can use the information as an aid in deciding how long to make the tests — or, if need be, in defending the test against a claim that it is not a reliable basis for making pass/fail decisions.

These output formats are illustrated in the handout. The first page shows a contingency table formed from the joint distribution by specifying a pass/fail cutoff score for observed scores and a similar cutoff for true scores. The percentages in each cell are estimated percentages of the population of exxaminees. For example, an estimated 11% of the examinees will pass the examination even though their true scores are below the cutoff.

At the bottom of the first page is an index labeled "decision-making efficiency". This index is based on an assumption of linear utility.

Linear utility means that if Jones has a true score ten points above the cutoff, and Smith has a true score five points above the cutoff, then failing Jones is twice as serious an error as failing Smith. The efficiency index is simply the utility of selecting on observed scores, divided by the utility of selecting on true scores. Method 20 makes it possible to estimate such an index.

The second page of the handout shows the estimated distribution of true scores in the passing group and in the failing group. This table gives the test user a good idea of the amount of separation between the two groups. If the test were perfectly reliable there would be no overlap between the true-score distributions of the two groups. If the test were completely unreliable, the two true-score distributions would be identical.

The third page of the handout shows the estimated distribution of true scores for examinees having a specified observed score. (This particular table is for examinees having the lowest passing score.) The more reliable the test, the less the spread in this distribution.

The fourth page of the handout shows a series of predicted observed-score distributions for examinees having specified true scores. If the user does not specify the levels, the program will print the distributions for true scores of .10, .30, .50, .70, and .90. These distributions are only slightly different from the binomial distributions for n = 30 (the number of items on the test).

The next four pages of the handout are the same as the first four, except that the data from the thirty-item test has been used to estimate distributions for a sixty-item test. By comparing the corresponding tables, you can see how much improvement would result from making the test twice as long. The last two pages of the handout are

computer-drawn graphs corresponding to the tables on pages 6 and 7 of the handout. The graphs enable the user of the program to <u>see</u> the overlap in the true-score distributions for the passing and failing groups and the spread in the true-score distribution for examinees with a given observed score. (The numbers on the vertical scale in these graphs do not correspond to those in the tables, because the true-score scale has been divided into intervals of .01 instead of .05.)

What are the implications of this type of reliability estimation? I believe that these applications of Lord's work will change the ways that we in the testing profession think about reliability, at least for tests that are intended to be the basis for pass/fail decisions. For these types of tests, I think we would be justified in saying that the traditional measures of reliability - the reliability coefficient and the standard error of measurement - are now or will soon be obsolete.

REFERENCES

Lord, F. M. Estimating true-score distributions in psychological testing (an empirical Bayes estimation problem). Psychometrika, 1969, 34, 259-299.

Wingersky, M.S., Lees, D.M., Lennon, V., & Lord, F. M. A computer program for estimating true-score distributions and graduating observed-score distributions. Educational and Psychological Measurement, 1969, 29, 689-692.

Samuel A. Livingston

Handout: page 1

Printed in U.S.A.

Reliability of Tests Used to Make Pass/Fail Decisions: Asking the Right Questions

NCME Annual Meeting: Toronto, Ontario; March, 1978

PASS-FAIL CONTINGENCY TABLE

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THIS TABLE IS FOR A TEST OF 30 ITEMS

TRUE-SCORE CUT-OFF : 0.6667

MINIMUM PASSING SCORE IS 20 OUT OF 30 ITEMS (67 PERCENT)

	SHOULD PASS	SHOULD FAIL
WILL PASS	55.9%	11.0%
WILL FAIL	8.8%	24.3%

DECISION-MAKING EFFICIENCY : 0.81

DECISION-MAKING EFFICIENCY IS DEFINED AS

$$\frac{\text{EXPECTATION}((Z-Z^*)\ \text{SIGN}(X-X^*))}{\text{EXPECTATION}((Z-Z^*)\ \text{SIGN}(Z-Z^*))}$$

WHERE Z* IS TRUE-SCORE CUT-OFF OF 0.67
X* IS OBSERVED-SCORE CUT-OFF OF 20
(MINIMUM PASSING SCORE MINUS ONE-HALF ITEM)

CONDITIONAL TRUE-SCORE DISTRIBUTIONS FOR PASSING AND FAILING GROUPS.

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THESE DISTRIBUTIONS ARE FOR A TEST OF 30 ITEMS, WITH A MINIMUM PASSING SCORE OF 20 ITEMS CORRECT

TRUE-SCORE (PROPORTION-CORRECT)	NUMBER PASSING GROUP	NUMBER FAILING GROUP	PERCENT PASSING GROUP	PERCENT FAILING GROUP	CUMULATIVE PERCENT PASSING GROUP	CUMULATIVE PERCENT FAILING GROUP
0.95- 1.00	0.	0.	0.0	0.0	100.0	100.0
0.90- 0.95	20.	0.	0.9	0.0	100.0	100.0
0.85- 0.90	115.	0.	5.2	0.0	99.1	100.0
0.80- 0.85	309.	2.	14.1	0.2	93.8	100.0
0.75- 0.80	521.	27.	23.8	2.5	79.8	99.8
0.70- 0.75	570.	113.	26.0	10.4	56.0	97.3
0.65- 0.70	405.	236.	18.5	21.8	30.0	86.9
0.60- 0.65	186.	285.	8.5	26.3	11.5	65.1
0.55- 0.60	55.	223.	2.5	20.6	3.0	38.8
0.50- 0.55	10.	123.	0.5	11.3	0.5	18.2
0.45- 0.50	1.	50.	0.1	4.6	0.1	6.8
0.40- 0.45	0.	15.	0.0	1.4	0.0	2.3
0.35- 0.40	0.	4.	0.0	0.3	0.0	0.9
0.30- 0.35	0.	1.	0.0	0.1	0.0	0.5
0.25- 0.30	0.	1.	0.0	0.1	0.0	0.4
0.20- 0.25	0.	1.	0.0	0.1	0.0	0.3
0.15- 0.20	0.	1.	0.0	0.1	0.0	0.3
0.10- 0.15	0.	1.	0.0	0.1	0.0	0.2
0.05- 0.10	0.	1.	0.0	0.1	0.0	0.1
0.0 - 0.05	0.	1.	0.0	0.1	0.0	0.1
TRUE-SCORE MEAN	0.74	0.62				
S.D.	0.073	0.080				

CONDITIONAL TRUE-SCORE DISTRIBUTION FOR SINGLE OBSERVED-SCORE LEVEL

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THIS DISTRIBUTION IS FOR A TEST OF 30 ITEMS

FOR EXAMINEES WITH OBSERVED-SCORE 20 OUT OF 30 ITEMS (67 PERCENT)

TRUE-SCORE (PROPORTION-CORRECT)	NUMBER	PERCENT	CUMULATIVE PERCENT
0.95- 1.00	0.	0.0	100.0
0.90- 0.95	0.	0.0	100.0
0.85- 0.90	0.	0.1	100.0
0.80- 0.85	5.	1.5	100.0
0.75- 0.80	33.	10.1	99.9
0.70- 0.75	85.	26.0	98.4
0.65- 0.70	103.	31.6	86.5
0.60- 0.65	66.	20.2	62.3
0.55- 0.60	26.	7.9	30.7
0.50- 0.55	6.	1.8	9.9
0.45- 0.50	1.	0.3	2.0
0.40- 0.45	0.	0.0	0.3
0.35- 0.40	0.	0.0	0.0
0.30- 0.35	0.	0.0	0.0
0.25- 0.30	0.	0.0	0.0
0.20- 0.25	0.	0.0	0.0
0.15- 0.20	0.	0.0	0.0
0.10- 0.15	[illegible]	0.0	0.0
0.05- 0.10	[illegible]	0.0	0.0
0.00- 0.05	[illegible]	0.0	0.0

TRUE-SCORE
MEAN 0.68
S.D. 0.060

PREDICTED CONDITIONAL OBSERVED-SCORE DISTRIBUTIONS FOR SELECTED TRUE-SCORE LEVELS

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THESE DISTRIBUTIONS ARE FOR A TEST OF 30 ITEMS

	NON-CUMULATIVE DISTRIBUTIONS					CUMULATIVE DISTRIBUTIONS				
	TRUE-SCORE (PROPORTION-CORRECT)					TRUE-SCORE (PROPORTION-CORRECT)				
OBSERVED-SCORE	.10	.30	.50	.70	.90	.10	.30	.50	.70	.90
30 - 30	0.0	0.0	0.0	0.0	3.2	100.0	100.0	100.0	100.0	100.0
29 - 29	0.0	0.0	0.0	0.0	13.0	100.0	100.0	100.0	100.0	96.8
28 - 28	0.0	0.0	0.0	0.1	23.4	100.0	100.0	100.0	100.0	83.8
27 - 27	0.0	0.0	0.0	0.5	25.5	100.0	100.0	100.0	99.9	60.4
26 - 26	0.0	0.0	0.0	1.6	18.9	100.0	100.0	100.0	99.4	35.0
25 - 25	0.0	0.0	0.0	4.0	10.2	100.0	100.0	100.0	97.8	16.0
24 - 24	0.0	0.0	0.0	7.9	4.1	100.0	100.0	100.0	93.8	5.8
23 - 23	0.0	0.0	0.1	12.4	1.3	100.0	100.0	100.0	86.0	1.7
22 - 22	0.0	0.0	0.3	15.9	0.3	100.0	100.0	100.0	73.6	0.4
21 - 21	0.0	0.0	1.0	16.9	0.1	100.0	100.0	99.9	57.7	0.1
20 - 20	0.0	0.0	2.3	15.1	0.0	100.0	100.0	99.5	40.7	0.0
19 - 19	0.0	0.0	4.6	11.4	0.0	100.0	100.0	98.6	25.6	0.0
18 - 18	0.0	0.0	7.9	7.3	0.0	100.0	100.0	96.3	14.2	0.0
17 - 17	0.0	0.1	11.5	4.0	0.0	100.0	100.0	91.6	6.9	0.0
16 - 16	0.0	0.2	14.4	1.8	0.0	100.0	100.0	83.7	2.9	0.0
15 - 15	0.0	0.7	15.6	0.7	0.0	100.0	99.9	72.2	1.1	0.0
14 - 14	0.0	1.8	14.4	0.2	0.0	100.0	99.7	57.8	0.3	0.0
13 - 13	0.0	4.0	11.5	0.1	0.0	100.0	98.9	42.2	0.1	0.0
12 - 12	0.0	7.3	7.9	0.0	0.0	100.0	97.1	27.8	0.0	0.0
11 - 11	0.0	11.4	4.6	0.0	0.0	100.0	93.1	16.3	0.0	0.0
10 - 10	0.0	15.1	2.3	0.0	0.0	100.0	85.8	8.4	0.0	0.0
9 - 9	0.1	16.9	1.0	0.0	0.0	100.0	74.4	3.7	0.0	0.0
8 - 8	0.3	15.9	0.3	0.0	0.0	99.9	59.3	1.4	0.0	0.0
7 - 7	1.3	12.4	0.1	0.0	0.0	99.6	42.3	0.5	0.0	0.0
6 - 6	4.1	7.9	0.0	0.0	0.0	98.3	26.4	0.1	0.0	0.0
5 - 5	10.2	4.0	0.0	0.0	0.0	94.2	14.0	0.0	0.0	0.0
4 - 4	18.9	1.6	0.0	0.0	0.0	84.0	6.2	0.0	0.0	0.0
3 - 3	25.5	0.5	0.0	0.0	0.0	65.0	2.2	0.0	0.0	0.0
2 - 2	23.4	0.1	0.0	0.0	0.0	39.6	0.6	0.0	0.0	0.0
1 - 1	13.0	0.0	0.0	0.0	0.0	16.2	0.1	0.0	0.0	0.0
0 - 0	3.2	0.0	0.0	0.0	0.0	3.2	0.0	0.0	0.0	0.0

PASS-FAIL CONTINGENCY TABLE

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THIS TABLE IS FOR A TEST OF 60 ITEMS

TRUE-SCORE CUT-OFF : 0.6667

MINIMUM PASSING SCORE IS 40 OUT OF 60 ITEMS (67 PERCENT)

	SHOULD PASS	SHOULD FAIL
WILL PASS	57.6%	8.2%
WILL FAIL	7.1%	27.0%

DECISION-MAKING EFFICIENCY : 0.86

DECISION-MAKING EFFICIENCY IS DEFINED AS $\frac{\text{EXPECTATION}((Z-Z^*)\ \text{SIGN}(X-X^*))}{\text{EXPECTATION}((Z-Z^*)\ \text{SIGN}(Z-Z^*))}$

WHERE Z* IS TRUE-SCORE CUT-OFF OF 0.67
X* IS OBSERVED-SCORE CUT-OFF OF 40
(MINIMUM PASSING SCORE MINUS ONE-HALF ITEM)

CONDITIONAL TRUE-SCORE DISTRIBUTIONS FOR PASSING AND FAILING GROUPS

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THESE DISTRIBUTIONS ARE FOR A TEST OF 60 ITEMS, WITH A MINIMUM PASSING SCORE OF 40 ITEMS CORRECT

TRUE-SCORE (PROPORTION-CORRECT)	NUMBER PASSING GROUP	NUMBER FAILING GROUP	PERCENT PASSING GROUP	PERCENT FAILING GROUP	CUMULATIVE PERCENT PASSING GROUP	CUMULATIVE PERCENT FAILING GROUP
0.95- 1.00	0.	0.	0.0	0.0	100.0	100.0
0.90- 0.95	20.	0.	0.9	0.0	100.0	100.0
0.85- 0.90	115.	0.	5.3	0.0	99.1	100.0
0.80- 0.85	311.	0.	14.4	0.0	93.7	100.0
0.75- 0.80	538.	9.	25.0	0.8	79.3	100.0
0.70- 0.75	603.	80.	28.0	7.2	54.4	99.2
0.65- 0.70	399.	241.	18.5	21.6	26.4	92.0
0.60- 0.65	142.	329.	6.6	29.4	7.9	70.4
0.55- 0.60	25.	253.	1.2	22.6	1.3	41.0
0.50- 0.55	2.	131.	0.1	11.7	0.1	18.4
0.45- 0.50	0.	51.	0.0	4.8	0.0	6.7
0.40- 0.45	0.	15.	0.0	1.4	0.0	2.2
0.35- 0.40	0.	4.	0.0	0.3	0.0	0.8
0.30- 0.35	0.	1.	0.0	0.1	0.0	0.5
0.25- 0.30	0.	1.	0.0	0.1	0.0	0.4
0.20- 0.25	0.	1.	0.0	0.1	0.0	0.3
0.15- 0.20	0.	1.	0.0	0.1	0.0	0.3
0.10- 0.15	0.	1.	0.0	0.1	0.0	0.2
0.05- 0.10	0.	1.	0.0	0.1	0.0	0.1
0.0 - 0.05	0.	1.	0.0	0.1	0.0	0.1
TRUE-SCORE MEAN	0.74	0.61				
S.D.	0.067	0.074				

CONDITIONAL TRUE-SCORE DISTRIBUTION FOR SINGLE OBSERVED-SCORE LEVEL

NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THIS DISTRIBUTION IS FOR A TEST OF 60 ITEMS

FOR EXAMINEES WITH OBSERVED-SCORE 40 OUT OF 60 ITEMS (67 PERCENT)

TRUE-SCORE (PROPORTION-CORRECT)	NUMBER	PERCENT	CUMULATIVE PERCENT
0.95- 1.00	0.	0.0	100.0
0.90- 0.95	0.	0.0	100.0
0.85- 0.90	0.	0.0	100.0
0.80- 0.85	0.	0.2	100.0
0.75- 0.80	9.	5.0	99.8
0.70- 0.75	46.	25.2	94.7
0.65- 0.70	72.	39.5	69.5
0.60- 0.65	43.	23.5	30.0
0.55- 0.60	11.	5.8	6.5
0.50- 0.55	1.	0.6	0.7
0.45- 0.50	0.	0.0	0.0
0.40- 0.45	0.	0.0	0.0
0.35- 0.40	0.	0.0	0.0
0.30- 0.35	0.	0.0	0.0
0.25- 0.30	0.	0.0	0.0
0.20- 0.25	0.	0.0	0.0
0.15- 0.20	0.	0.0	0.0
0.10- 0.15	0.	0.0	0.0
0.05- 0.10	0.	0.0	0.0
0.00- 0.05	0.	0.0	0.0

TRUE-SCORE	
MEAN	0.67
S.D.	0.048

PREDICTED CONDITIONAL OBSERVED-SCORE DISTRIBUTIONS FOR SELECTED TRUE-SCORE LEVELS

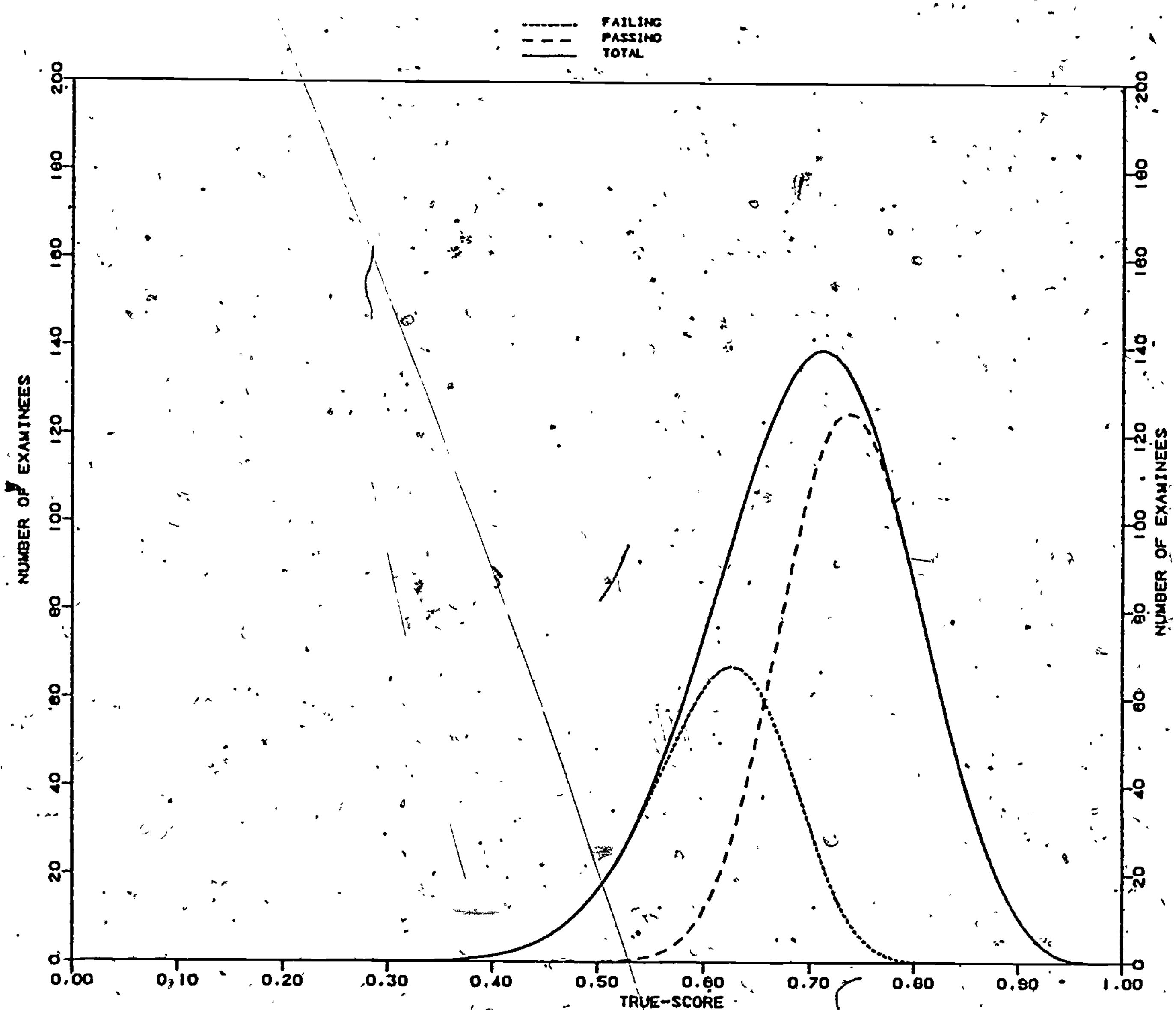
NAME OF TEST :

ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30

THESE DISTRIBUTIONS ARE FOR A TEST OF 60 ITEMS

	NON-CUMULATIVE DISTRIBUTIONS TRUE-SCORE (PROPORTION-CORRECT)					CUMULATIVE DISTRIBUTIONS TRUE-SCORE (PROPORTION-CORRECT)				
OBSERVED-SCORE	.10	.30	.50	.70	.90	.10	.30	.50	.70	.90
58 - 60	0.0	0.0	0.0	0.0	4.1	100.0	100.0	100.0	100.0	100.0
55 - 57	0.0	0.0	0.0	0.0	38.5	100.0	100.0	100.0	100.0	95.9
52 - 54	0.0	0.0	0.0	0.1	44.9	100.0	100.0	100.0	100.0	57.4
49 - 51	0.0	0.0	0.0	2.0	11.6	100.0	100.0	100.0	99.9	12.5
46 - 48	0.0	0.0	0.0	12.2	0.9	100.0	100.0	100.0	97.9	0.9
43 - 45	0.0	0.0	0.0	30.2	0.0	100.0	100.0	100.0	85.6	0.0
40 - 42	0.0	0.0	0.4	33.2	0.0	100.0	100.0	100.0	55.4	0.0
37 - 39	0.0	0.0	3.1	17.3	0.0	100.0	100.0	99.6	22.2	0.0
34 - 36	0.0	0.0	13.0	4.4	0.0	100.0	100.0	96.5	5.0	0.0
31 - 33	0.0	0.0	28.0	0.6	0.0	100.0	100.0	83.5	0.6	0.0
28 - 30	0.0	0.2	31.2	0.0	0.0	100.0	100.0	55.5	0.0	0.0
25 - 27	0.0	2.4	18.1	0.0	0.0	100.0	99.7	24.4	0.0	0.0
22 - 24	0.0	11.8	5.4	0.0	0.0	100.0	97.4	6.3	0.0	0.0
19 - 21	0.0	29.0	0.8	0.0	0.0	100.0	85.6	0.9	0.0	0.0
16 - 18	0.0	34.0	0.1	0.0	0.0	100.0	56.6	0.1	0.0	0.0
13 - 15	0.3	18.2	0.0	0.0	0.0	100.0	22.6	0.0	0.0	0.0
10 - 12	5.6	4.1	0.0	0.0	0.0	99.7	4.4	0.0	0.0	0.0
7 - 9	33.3	0.3	0.0	0.0	0.0	94.2	0.3	0.0	0.0	0.0
4 - 6	49.0	0.0	0.0	0.0	0.0	60.8	0.0	0.0	0.0	0.0
1 - 3	11.8	0.0	0.0	0.0	0.0	11.9	0.0	0.0	0.0	0.0
0 - 0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0

CONDITIONAL TRUE-SCORE DISTRIBUTIONS FOR PASSING AND FAILING GROUPS AND TRUE-SCORE DISTRIBUTION
NAME OF TEST :
ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30
THESE DISTRIBUTIONS ARE FOR A TEST OF 60 ITEMS, WITH A MINIMUM PASSING SCORE OF 40 ITEMS CORRECT



CONDITIONAL TRUE-SCORE DISTRIBUTION FOR A SINGLE OBSERVED-SCORE
NAME OF TEST :
ORIGINAL DATA : ADMINISTRATION DATE = 07/01/77 NUMBER OF EXAMINEES = 3274 NUMBER OF ITEMS = 30
THIS DISTRIBUTION IS FOR A TEST OF 60 ITEMS
PERCENT OF EXAMINEES WITH OBSERVED-SCORE 40 OUT OF 60 ITEMS (67 PERCENT)